Explore Weather Trend

Linda Xu

12.22.2020

Outline：

In this project, I will analyze local and global temperature data and compare the temperature trends where I live to overall global temperature trends. I choose local area Anshan, China. And I'll be using SQL, EXCEL and R.

Step 1: Extraction of data from database using SQL

select * from global_data;

select * from city_data where country='China' and city='Anshan;

I download files as "results(1).csv" and "results(2).csv".

And I manipulate two files into one, name "New update.csv"

Moving Average:

Moving averages are used to smooth out data to make it easier to observe long term trends and not get lost in daily fluctuations. I create two columns called global mov_avg and local mov_avg, which are where the moving average field will be stored. I use commend =AVERAGE(B2:B11) for global mov_avg and commend=AVERAGE(E81:E90) for local mov_avg to see the moving average value for 10 years. Note that global temperature was recorded since 1750 while local temperature was recorded since 1829.

```
# Use ggplot2 for visualization
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.0.2

library(tidyverse)

## -- Attaching packages ---------------------------------------------------
------------ tidyverse 1.3.0 --

## v tibble  3.0.1     v dplyr   1.0.0
## v tidyr   1.1.0     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0
## v purrr   0.3.4

## Warning: package 'stringr' was built under R version 4.0.2

## -- Conflicts ------------------------------------------------------------
------- tidyverse_conflicts() --
```

```
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

Step 2: Read data and Visualization：

```
temp <- read.csv(file = 'C:/Users/xuwen/Desktop/Udecity/Project 1/New update.
csv')
head(temp)
```

```
##   year.1 Global.avg_temp global.mov_avg year.2 Anshan.avg_temp local.mov_a
vg
## 1   1750            8.72             NA     NA               NA
NA
## 2   1751            7.98             NA     NA               NA
NA
## 3   1752            5.78             NA     NA               NA
NA
## 4   1753            8.39             NA     NA               NA
NA
## 5   1754            8.47             NA     NA               NA
NA
## 6   1755            8.36             NA     NA               NA
NA
```

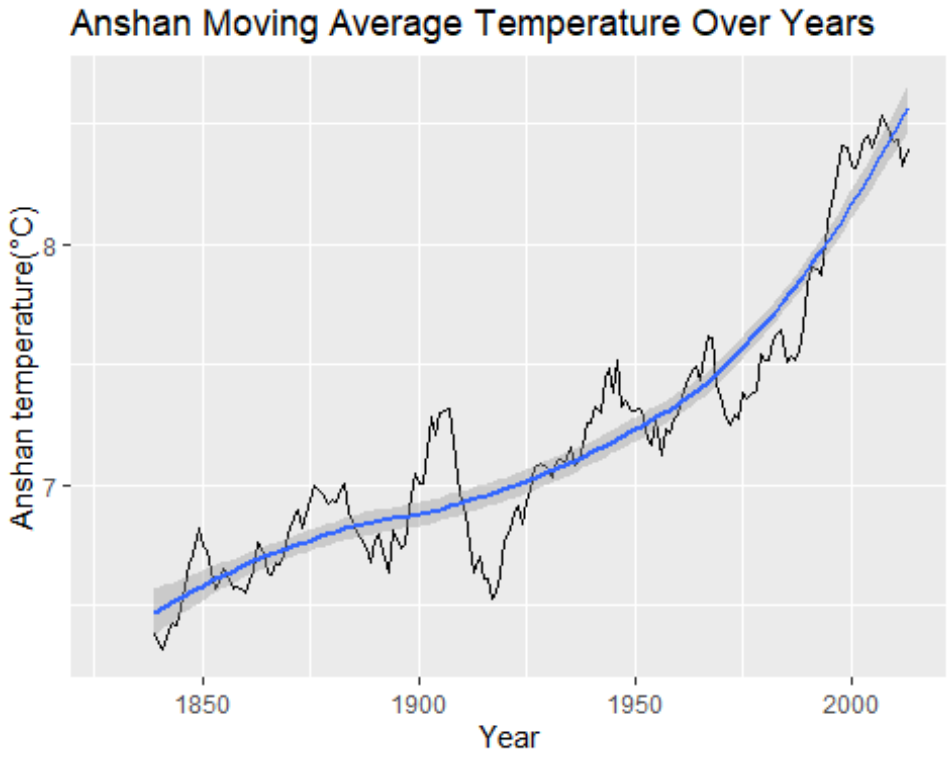
```
#start year 1829 for both local and global temperature
temp2<- read.csv(file = 'C:/Users/xuwen/Desktop/Udecity/Project 1/New update
2.csv')
head(temp2)
```

```
##   year.1 Global.avg_temp global.mov_avg year.2 Anshan.avg_temp local.mov_a
vg
## 1   1829            7.94          8.184   1829           11.80
NA
## 2   1830            8.52          8.274   1830            7.12
NA
## 3   1831            7.64          8.229   1831            6.02
NA
## 4   1832            7.45          8.155   1832            6.00
NA
## 5   1833            8.01          8.184   1833            6.59
NA
## 6   1834            8.15          8.144   1834            6.94
NA
```

Anshan Moving Average Temperature

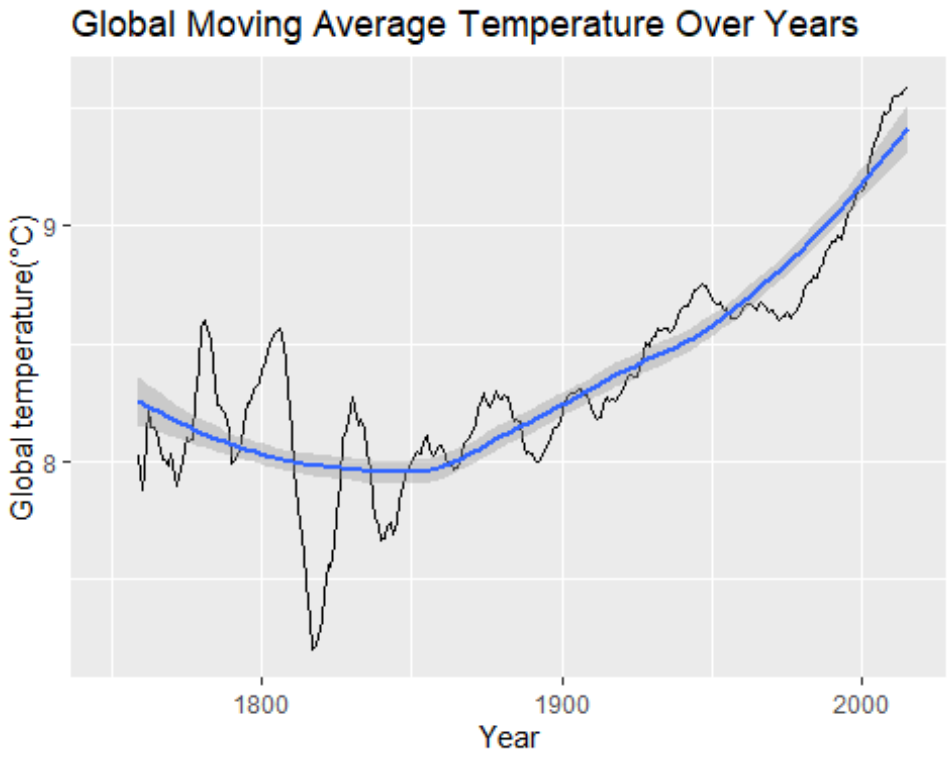
```
temp %>%
  select(year.2, local.mov_avg)%>%
  group_by(year.2)%>%
  ggplot(aes(y= local.mov_avg, x= year.2))+ geom_line()+geom_smooth()+ ggtitl
```

```
e("Anshan Moving Average Temperature Over Years")+labs(y="Anshan temperature(
°C)", x = "Year")

## `geom_smooth()` using method = 'loess' and formula 'y ~ x'

## Warning: Removed 91 rows containing non-finite values (stat_smooth).

## Warning: Removed 91 row(s) containing missing values (geom_path).
```



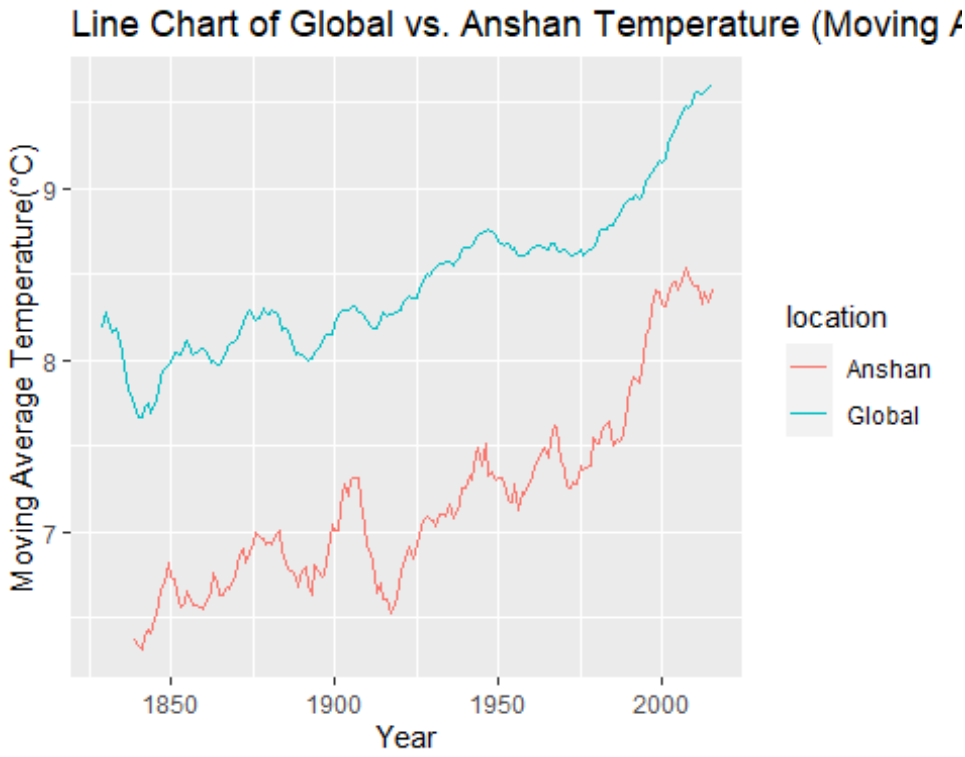
Global Moving Average Temperature

```
temp %>%
  select(year.1, global.mov_avg)%>%
  group_by(year.1)%>%
  ggplot(aes(y= global.mov_avg, x= year.1))+ geom_line()+geom_smooth()+ ggtit
le("Global Moving Average Temperature Over Years")+labs(y="Global temperature
(°C)", x = "Year")

## `geom_smooth()` using method = 'loess' and formula 'y ~ x'

## Warning: Removed 9 rows containing non-finite values (stat_smooth).

## Warning: Removed 9 row(s) containing missing values (geom_path).
```



Anshan and Global Moving Average Temperature

```
ggplot(temp2, aes(year.1)) +
  geom_line(aes(y = local.mov_avg, color = "blue")) +
  geom_line(aes(y = global.mov_avg, color= "yellow"))+
ggtitle("Line Chart of Global vs. Anshan Temperature (Moving Average) ")+labs
(y="Moving Average Temperature(°C)", x = "Year")+scale_color_discrete(name =
"location", labels = c("Anshan", "Global"))

## Warning: Removed 10 row(s) containing missing values (geom_path).
```




```
summary(temp2)
```

```
##      year.1     Global.avg_temp global.mov_avg      year.2     Anshan.avg_
temp
##  Min.   :1829   Min.   :7.380   Min.   :7.666   Min.   :1829   Min.   : 5.
450
##  1st Qu.:1876   1st Qu.:8.130   1st Qu.:8.139   1st Qu.:1875   1st Qu.: 6.
728
##  Median :1922   Median :8.500   Median :8.356   Median :1921   Median : 7.
130
##  Mean   :1922   Mean   :8.492   Mean   :8.461   Mean   :1921   Mean   : 7.
226
##  3rd Qu.:1968   3rd Qu.:8.755   3rd Qu.:8.674   3rd Qu.:1967   3rd Qu.: 7.
628
##  Max.   :2015   Max.   :9.830   Max.   :9.594   Max.   :2013   Max.   :11.
800
##                                                 NA's   :2      NA's   :3

##  local.mov_avg
##  Min.   :6.314
##  1st Qu.:6.760
##  Median :7.101
##  Mean   :7.200
##  3rd Qu.:7.439
##  Max.   :8.536
##  NA's   :10
```

Step 3: Observation

1. The temperature is rising over the years due to climate change.

2. The temperature is more than 2 degree higher between 1800 to 2100 on global basis.

3. The temperature in Anshan, China is lower than the global scale, but its temperature has increased a lot as well. Both of them are rising.

4. In early 1900, in terms of the Anshan moving average temperature, it was decreased for estimate 20 years, then it went up.

5. The lowest global moving average temperature is 7.67, and highest is 9.594; while the lowest Anshan moving average temperature is 6.314, and highest is 8.536. And the mean global moving average temperature is 8.461, while the mean local moving average temperature is 7.20.